Predicting Biological Cleanliness: An Empirical Bayes Approach for Spacecraft Bioburden Accounting

J. Nick Benardini
Jet Propulsion Laboratory,
California Institute of Technology
4800 Oak Grove Dr.
Pasadena, CA 91109
James.N.Benardini@jpl.nasa.gov

Arman Seuylemezian
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Dr.
Pasadena, CA 91109
Arman.Seuylemezian@jpl.nasa.gov

Andrei Gribok
Idaho National Laboratory
2525 N. Fremont Avenue
P.O. Box 1625
Idaho Falls, Idaho 83415
Andrei.Gribok@inl.gov

Abstract — To comply with the international planetary protection (PP) policy set forth by the Committee on Space Research (COSPAR) and National Aeronautics and Space Administration (NASA) Agency level requirements, spacecraft destined to biologically sensitive planetary bodies should "conduct exploration of them so as to avoid their harmful contamination". Analysis, testing, and inspection are the standard forward verification activities used to demonstrate compliance with biological contamination requirements. For testing of spacecraft surface areas, a swab or wipe sample is collected from surfaces prior to last access and subsequently processed in the lab using NASA-approved PP methods for culture-based assays. Raw data resulting from this assay is then statistically treated employing a mathematical paradigm stemming from the Mariner Mars 1971 Project to generate the bioburden density and total microbial bioburden present. This standard approach arbitrarily accounts for error and provides an upper conservative bound as it reports the maximum number of spores estimated to be present on flight hardware surfaces. A bioburden density estimates factors in the following variables: (1) observed bioburden count; (2) representative volume processed; (3) sampling device efficiencies; and (4) sampled surface area. Notably, to account for potential errors in the approach, a 0 observed count is changed to a NASA policy derived count of 1.

The data generated by spacecraft bioburden verification campaigns in the past have resulted in >80% of wipes and >90% of swabs containing a bioburden count of 0. As such, having a robust and well documented statistical approach for dealing with the probability of low incident rates is necessary to be able to estimate spacecraft bioburden. Being able to statistically describe the bioburden distribution and associated confidence level is a game-changer for the development of bioburden allocations during mission design and will allow for tighter management of risk throughout spacecraft build. Thus, employing an empirical Bayes (EB) statistical approach was evaluated to estimate the microbial bioburden on spacecraft to mitigate the aforementioned mathematical concerns and provide a probabilistic bioburden distribution of flight hardware surfaces.

For application of this approach to performing bioburden calculations, a range of non-informative prior assumptions on hardware surfaces are explored for Bayesian analyses while informative priors using posterior distributions from prior assays are utilized for EB analyses. Several non-informative priors are currently under investigation to assess fitness, including the use of a non-informative prior distribution bounded by the currently utilized NASA specification values for a basis of risk to account for unknowns during the spacecraft integration and test process. Informative priors under consideration are generated using sampled bioburden values from hardware originating within like processing environments (e.g., vendor cleaning process or similar assembly process), temporal spacecraft status events as a prediction for hardware cleanliness of future samples, and heritage system bioburden actuals to predict allocation for subsequent missions. Informative priors and probabilistic bioburden distributions are then validated using data sets from the Mars Exploration Rover, Mars Science Laboratory, and InSight missions. Using the EB approach to generate a probabilistic bioburden distribution as demonstrated through mission use cases provides a valid approach for use in the end-to-end requirements verification process.

978-1-7821-2734-7/20/$31.00 ©2020 IEEE

TABLE OF CONTENTS

1. INTRODUCTION

Planetary protection (PP) is a discipline that focuses on minimizing the biological contamination of spacecraft to ensure compliance with international policy. The National Aeronautics and Space Administration (NASA) has developed a set of requirements (NPR 8020.12) based on recommendations from the Committee on Space Research (COSPAR) that each mission must comply with regarding both forward and backward PP. Forward PP addresses the risk that an outbound spacecraft may carry Earth-based material (both organisms and biogenic compounds) and result in a contamination event on a celestial body. Backward PP focuses on preventing potential adverse effects to Earth's biosphere from returning unsterilized extraterrestrial particles. Forward PP requirements increase in stringency for spacecraft destined to a solar system body that is of astrobiological interest for life detection studies (either extant

or extinct) and/or the spacecraft's potential to introduce terrestrial life to a celestial body.

Biological cleanliness requirements to target bodies, such as Mars, include spacecraft assembly control and direct testing of the microbial bioburden to maintain spore requirements of 5×10^5 spores at launch with an average of 300 spores/m^2 bioburden density on flight hardware surfaces, while preventing recontamination by utilizing International Organization for Standardization (ISO) 8 or better cleanroom environments. Direct flight hardware testing is conducted using the NASA standard spore assay (NASA HBK 6022). Briefly, swabs or wipes are used to recover biologicals present on a given hardware component; samples are then taken back to the microbiology lab where the biologicals are removed from the sampling matrices using sonication and suspended in a buffered solution; and then the samples are heat shocked at 80°C +/-2°C for 15 minutes to select for spores, and subsequently grown in oligotrophic, aerobic conditions for 72 hours to assess the presence of colony forming units (CFU). These raw CFU counts are mathematically treated to generate a bioburden density and subsequently assigned to their respective hardware components outlined in the PP Equipment List (PPEL) of the project. Then, respective surface areas and volumes of each hardware component are taken into account to perform roll-up calculations based on the hardware hierarchy to generate a final spacecraft level bioburden estimate.

For Mars-bound spacecraft, a statistical paradigm stemming from the Mariner Mars 1971 mission was implemented to calculate a bioburden density from raw CFU counts. The 1971 mission used raw CFU counts from microbial verification assays that were mathematically processed using a sum-of-the-means approach in the following manner with the variables defined in Table 1. The approach for estimation of microbial bioburden on spacecraft was reported for mission use and not referenced or further detailed as to why this approach was selected.

Table 1. Definition of Variables Used in Bioburden Calculations

Variable	Definition
A_0	the total area represented by a group or sample set, m^2
n_s, n_w	the total number of swabs or wipes
n_{tot}	the total number of samples
a_{si}, a_{wj}	is the area sampled by the i^{th} swab and the j^{th} wipe, (m^2)
f_s, f_w	the pour fractions for swabs and wipes
e_s, e_w	the recovery efficiencies for swabs and wipes
A_S	the total effective area, (m^2)
N_{si}	the number of CFU counted in the i^{th} swab sample
N_{wj}	the number of CFU counted in the j^{th} wipe sample
N_{tot}	the total number of CFU in a group
B	the bioburden density is the total number of spores divided by the total effective area sampled, N_{tot}/A_s, (spores/m^2)

The total effective area sampled was represented by:

$$\boldsymbol{A_s = \sum_{i=1}^{n_s} a_{s_i} f_{s_i} e_{s_i}} \quad (1)$$

The total number of spores counted was:

$$\boldsymbol{N_{tot} = \sum_{i=1}^{n_s} N_{s_i}} \quad (2)$$

The bioburden density, B, was given by:

$$\mathbf{B = N_{tot} / A_s} \quad (3)$$

The estimate of the total bioburden, N, was given by:

$$\mathbf{N = B\ A_0} \quad (4)$$

The swab efficiency corrective factor utilized for the Mariner Mars 1971 project was intended to account for the efficiency of the sampling device to both collect biologicals from the sample surface and subsequently release the biologicals from the sample device for analysis. The experimental factor was derived from an internal report by Angelotti et al. (1964) entitled, "Comparative evaluation of the cotton swab and Rodac methods for the recovery of Bacillus subtilis spore contamination from stainless steel surfaces" [1], which observed an efficiency of ~47%. Although there was an experimentally derived value of 47%, a value of 30% was used for the mission as a "conservative factor," as was noted in the mission's June 29, 1970, Microbiological Assay and Monitoring Plan. No further rationale was noted as to why this conservative value was utilized.

For the Viking missions launched in 1975, a robotic spacecraft bioburden prediction model was developed and implemented into a management tool to simulate microbial contamination during spacecraft assembly and minimize the number of samples collected in order to predict the bioburden distribution prior to microbial reduction and launch [2]. These predictions were utilized during the spacecraft build to ultimately reduce the number of direct verification assays and engineering controls on the two spacecraft to a combined total of 4,899 swabs during the final eight months of integration and tests [3]. Notably, for the Viking missions there were only eight geographically distinct spacecraft groupings representing several hardware components derived from the thermal heating analysis performed for the terminal dry heat microbial reduction. The Viking missions swab collection and processing efficiency corrective factor that was utilized, as agreed upon by both NASA headquarters (HQ) and the mission, was 50%. No further explanation was documented in the Viking 1975 Project bioburden model report as to the selection of a 50% efficiency. This efficiency value was only 3% above the study-reported value of 47%. The Mars Pathfinder Mission (MPF), which launched in 1996 and was the next NASA mission to land a spacecraft on the surface of Mars, adapted the Viking sum-of-the-means approach, but did not account for the sampling device efficiency and instead applied Poisson and Gaussian statistics utilizing a 3-sigma limit (3σ) to report a maximum predicted bioburden (B_{max}), as discussed in the Mars Pathfinder PP Plan [4].

The total effective area sampled was represented by:

$$\boldsymbol{A_s = \sum_{i=1}^{n_s} a_{s_i} f_{s_i} + \sum_{j=1}^{n_w} a_{w_j} f_{w_j}} \quad (5)$$

The total number of spores counted was:

$$N_{tot} = \sum_{i=1}^{n_s} N_{s_i} + \sum_{j=1}^{n_w} N_{w_j} \quad (6)$$

The bioburden density, B, was given by:

$$B = N_{tot} / A_s \quad (7)$$

The estimate of the bioburden maximum, B_{max}, was based on: (1) Poisson statistics for strictly swab samples $N_{tot}<1$; (2) Gaussian statistics for strictly wipe samples $N_{tot}<1$; and (3) Gaussian statistics for a combination of wipe and swab samples $N_{tot}>1$, while factoring in the sum of the mean. The Mars Pathfinder mission was the first to document that the probability per sample of a positive count is small over many samples for a given A_0; thus, a Poisson statistical approach would be valid for swab samples. The standard deviation for a group of swabs with a sampled area of A_0 was calculated by the mean of its square root as follows:

$$\sigma = 1 \div \sqrt{(A_0 A_s)} \quad (8)$$

Despite the mean bioburden being zero for the swab group a NASA policy directive was employed to change the count to 1 spore and subsequently calculate the B_{max}, which is given by

$$B_{max} = 1/A_s + 3\sigma \quad (9)$$

Conversely, for groups of wipes where the raw bioburden CFU was $N_{tot}>1$, Gaussian statistics were used because less wipes would be collected over a given A_0, which resulted in a standard deviation for a group of wipes being calculated as:

$$\sigma = 1 \div A_s \quad (10)$$

For groups of wipes where $N_{tot}<1$, a NASA policy directive was employed to change the count of 1 spore to calculate the B_{max}, given by

$$B_{max} = 1/A_s + 3\sigma \quad (11)$$

When $N_{tot}>1$ for a combination of wipes and swabs, the sigma is calculated with the actual number of CFUs as follows:

$$\sigma = \sqrt{(N/A_s)} \quad (12)$$

For groups of wipes and swabs where $N_{tot}>1$, the formal mean is used to calculate the B_{max}, which is given by:

$$B_{max} = N/A_s + 3\sigma \quad (13)$$

The Mars Exploration Rover (MER) mission baselined their bioburden calculation approach in a similar fashion to the MPF by using three different statistical treatments dependent upon sample type and observed CFU in the laboratory. The Mars Science Laboratory (MSL) further expanded the bioburden statistical approach to treat wipes following a Poisson distribution, due to the larger surface area of the spacecraft as compared to MPF or MER [5].

The MSL mission had to take a pragmatic approach to the previously applied Gaussian approach from MER for the wipe samples because the accountable surface area for the MSL mission was 25% more than for MER, therefore requiring more wipe samples. The raw bioburden statistical treatment scenarios for MSL included: (1) a general case for wipes and swabs where $N_{tot}>1$; (2) a swab only case where $N_{tot}<1$; (3) wipes and swabs where $N_{tot}<1$; and (4) one swab or wipe sample where $N_{tot}>1$, which is further detailed in [6], using over 29 equations to calculate the total bioburden for the mission. In general, raw CFU counts were treated utilizing Poisson statistics for a collection of samples where $N_{tot}<1$, or if there was only one sample in a hardware group. All others employed Gaussian ones. In addition to expanding the statistical treatments, MSL also significantly changed the calculations by employing a weighted average approach based on the surface area, where the standard deviations of each sample were weighted by the corresponding surface area sampled prior to averaging. This resulted in a preferential weighting to raw CFU counts generated from wipe samples due to the larger surface area sampled, thereby lowering the overall standard deviations of the reported bioburden densities. Unlike MER, the MSL statistical approach used the actual spore count in this case to determine the bioburden density as opposed to adding a NASA policy derived count of 1 to a group where no spores were observed.

Towards the end of the MSL mission, the European Space Agency and the NASA Planetary Protection Office coordinated efforts to standardize the raw bioburden calculation approach. This resulted in future missions, such as InSight and Mars 2020, to revert back to the Viking sum-of-the-means approach using equations 1-4 [7, 8]. In addition to standardizing this approach, swab and wipe recovery efficiencies (e_s, e_w), as shown in Table 2, were adopted, as well as reinforcing the assignment of a NASA policy derived count of 1 CFU where the true CFU count was zero observed spores. For the InSight mission, a series of weighted averages were used to calculate the bioburden. Data from each unique sampling date was treated independently, even though it originated from sampling the same hardware component resulting in multiple groupings for a single hardware component. A post-mission analysis of the impact of artificially converting a 0 CFU observed value to 1 CFU resulted in 127 hardware groups being manipulated, which yielded an increase of 1.7% in the total bioburden.

Table 2. Swab and Wipe Combined Sample Collection and Processing Efficiency Values

Sampling Device	Processing Method	% Efficiency	Traceability
Puritan Cotton Swab 806-WC	NASA	31.7	ESA/NASA Study[1]
COPAN Polyester Swab, PE ATP	European Space Agency (ESA)	11.1	ESA/NASA Study
COPAN Flocked Nylon, FLOQSwab 522C	NASA	13.2	ESA/NASA Study
COPAN Flocked Nylon, FLOQSwab 522C	ESA	49	ESA/NASA Study
TexWipe 3211	NASA	30	NASA HQ Value

[1] Validation of swab assay method report. ESA Study Report. June, 27, 2013. TEC-QI/13-50. C. Moissl-Eichinger.

The final bioburden density values for each component are then imported into the mission's PPEL as a means to keep track of the bioburden requirements compliance. These densities and corresponding total bioburdens from the hardware groups are then cumulatively added into various bins, such as landed hardware, impacting hardware, subsystem hardware, bioregions (*i.e.,* hardware groupings that are exposed to a particular space environment condition such as UV or radiation), or the entire spacecraft. These roll up calculations are performed in the PPEL using a series of lookup equations. Ultimately, the total bioburden of the spacecraft is then reported out and compared to the 'at-launch' bioburden requirements for compliance or utilized as an input parameter into the inadvertent biological probabilistic risk assessment (PRA).

The PRA is a mature and influential technology that relies on two core methodologies—fault/event trees and statistical parameter estimation. The contamination probability event trees analysis utilizes Boolean logic to combine different paths to a contamination event. Once such paths are exhaustively enumerated, the elementary probability rules are applied to aggregate the probabilities of different contamination scenarios into a contamination event [9]. The performance of PP PRA depends critically on the accuracy of parameter estimates for the individual components, as they are merged through event trees analysis to produce the overall probability of a serious contamination event. As an input parameter, bio-contamination is the initial bioburden at launch, which accounts for the number of microorganisms present on the spacecraft [9]. Historically, two complimentary approaches are used in PRA to estimate bioburden densities for individual components—frequentist and Bayesian. We consider the two approaches complimentary because the Bayesian approach complements the frequentists' likelihood function with prior information in the form of a previous distribution against the parameter of interest. This paper compares and contrasts these two approaches used for bioburden calculations and evaluates their performance using data collected from the InSight mission.

While these numerical approaches to bioburden calculations have been widely accepted and used for bioburden validation on past missions, the increased complexity and sensitivity of missions that are in development and the need to mathematically evaluate and document the approach are critical for future mission success. Such bioburden density models, as opposed to the previously utilized bioburden maximum, will allow for mission planning and PP trades to be conducted during the design stages of a mission. Particularly in developing a sampling plan, associated spacecraft surface area coverage, the implementation of appropriate cleanroom class protocols (i.e., stringency of gowning and cleanroom protocols), and microbial reduction planning. These models will also aid the mission throughout the build and testing phase, as they will allow for tighter monitoring and prediction of bioburden throughout the course of a mission. In addition, they can assist NASA HQ in their verification oversight activities and mission certification of flight readiness. For missions requiring a PRA, like those proposing to return samples or those destined for outer planets, the application of this type of model and associated mission dataset will be critical in evaluating the distribution of the microbial bioburden present on spacecraft hardware and associated confidence intervals. The distribution and confidence level reporting being proposed provides a critical next step in spacecraft contamination monitoring as it provides a robust mathematical framework that can be used to technically describe the biological contamination from bounding cases to probable cases to help bridge the science and policy gap.

2. Empirical Bayes Bioburden Data Collection and Processing

The InSight mission PP verification dataset was used to develop an empirical Bayes (EB) bioburden estimation model. This dataset from 2013−2018 contained raw CFU counts from spacecraft status checks, hardware receiving inspection, hardware closeouts, spacecraft stack, and NASA verification assays of 2,031 swabs and 1,266 wipes. In general, this dataset had a low incidence of CFU's as 93% of the swabs and 63% of the wipes had a final CFU count of 0 at 72 hours, thereby resulting in ~85% of the 39,379 petri dishes yielding 0 CFU. Therefore, for the purposes of model development and the application of EB, a portion of the total InSight dataset was used containing strictly sampled test data as opposed to those components accounted for using specification values or a combination of test data and specification values. For the application of EB, test data from several components were used to generate informative priors, which were then applied to estimating the bioburden density of another component, thus demonstrating various engineering use cases for such a model.

The data for each individual spacecraft component were collected using either swabs or wipes. For each component, a number of samples were collected on a given date or on different dates. A swab data collection covered the area of 0.0025 m^2 with a single swab, while a wipe-covered area varied typically between 0.1 and 1.0 m^2, depending on the geometric complexity and size of the sampled component. Each swab or wipe was considered a sample. Having been processed in the microbiology laboratory, the samples were deposited onto petri dishes containing tryptic soy agar. For swabs, only 80% of aliquots were deposited in the dishes, thus producing a pour fraction of 0.8, which was taken into consideration by reducing the sampled area. For wipes, the pour fraction was 0.25. In this paper, the sampling efficiency of wipes or swabs have not been taken into account, as it is not a directly observable quantity and must be experimentally estimated. The number of samples was different for each component and each of them could have been sampled either with swabs, wipes, or both. For the purpose of Bayesian analysis, the raw data for each component was represented by pairs (x_i, e_i), i=1,2...N, where x_i is the number of CFU counts for i^{th} sample (swab or wipe) and e_i is the exposure calculated as the area covered with a swab or wipe multiplied by the corresponding pour ratio, and N is the number of samples

collected for a component. Data for each component were pooled to produce a total count and total exposure as $X=\sum_1^N x_i$ and $E=\sum_1^N e_i$. The total count and total exposure have been used in Poisson likelihood for Bayesian inference.

Sampled data from eight different InSight components were used in this paper. The components were selected to represent a wide range of raw CFU counts, sampled surface areas, exposures, and total surface areas. Exposure is utilized interchangeably with the term effective area throughout this paper. Table 3 summarizes the data used in this paper.

Table 3. Summary of Bioburden Data for Eight InSight Components

Component	CFU count	Area sampled, m^2	Exposure: area sampled × pour ratio, m^2	Total surface area of the component, m^2	% sampled=area sampled/total area
9	0	0.6031	0.2167	0.7580	79.5650
73	0	2.4200	0.6160	2.7400	88.3212
300	1	2.6600	0.6705	5.0000	53.2000
169	1	0.2400	0.1920	0.5850	41.0260
283	5	4.5710	1.1427	12.0000	38.0920
243	5	0.2800	0.1140	0.2980	93.9600
38	12	3.1050	0.8065	10.0000	31.0500
261	52	0.0600	0.0480	0.3120	19.2310

3 Constrained Non-Informative Priors: Utilization In PP Planning and Implementation

Priors are envisioned to inform the lifecycle of a flight project to include the planning, design, implementation, and monitoring of the spacecraft throughout the assembly, test, and launch operations phase. During the mission planning and design phases, empirical data from previous missions and hardware sets can be used to inform engineering judgement to develop the expected flow down of a mission's allotted bioburden. Beneficial use cases where an informative prior (generated from previous mission datasets) could be utilized to inform the current state-of-the-state of an active mission include a like-processing environment where the same cleanroom or cleaning process is employed allowing for the application of a unique informative prior for the respective vendor/hardware component, a particular microbial reduction process, or perhaps an entire heritage system (e.g., Phoenix prior being utilized for InSight). During the active build and test operations phase of the mission, informative priors generated from hardware receiving inspection or spacecraft health check assays can be applied to final assays to refine the current best estimates. In addition, the current relevant mission dataset can be used to generate informative priors to predict audit assays prior to launch, biological cleanliness before and after critical operations (e.g., spacecraft move, environmental test), and forecast bioburden prior to delivery of a given hardware component from a vendor's bioburden-controlled environment.

While the major appealing property of Bayesian inference is its ability to include prior information into the inference model, sometimes it is required to avoid reliance on old information as it may dominate the newly collected data, especially if the new data are sparse. With the development of missions to outer planets and the proposed Mars sample return campaign, a need exists to perform PP assessments on previously un-encountered spacecraft components and novel instruments. Implementing informative priors on these new mission architectures based on data from previous missions that may have different architectures, including new vendors or commercial entities—and thus a different expected bioburden—can inappropriately influence bioburden estimations. Another reason to do away with informative priors is the lack of reliable prior information about the parameter of interest. Finally, informative priors may be considered "subjective" by peers; as such, using noninformative priors may help to alleviate this concern. On the other hand, noninformative priors quite often could be improper leading to undefined moments or noninvariant situations under re-parameterization. Since the posterior mean is pulled toward the prior mean and the reciprocal of prior variance is the regularization parameter, noninformative priors can be quite defective.

This motivated the development of a constrained noninformative (CNI) prior [10], which while retaining desirable properties of a noninformative distribution, also allows specifying the expected value of the parameter of interest. The CNI uses a definition of entropy suggested in [11], which is the negative of the Kullback–Leibler divergence [12] with a reference distribution $\pi(\theta)$ being Jeffreys prior:

$$H = -\int p(\theta)\log\left(\frac{p(\theta)}{\pi(\theta)}\right) d\theta \tag{14}$$

In addition to the re-parametrization property, the CNI also has a larger variance than other noninformative priors, such as the maximum entropy prior [11]. This larger variance impacts the prior by diminishing its influence and gives more weight to the data. The variance of the prior distribution plays a prominent role in Bayesian inference as it defines how strongly the posterior mean is "pulled" towards the prior mean. The CNI can be parametrized as Gamma $(1/2,1/(2\cdot\mu))$, where μ is the pre-defined expected value. The mean of the posterior distribution has been reported as summary statistics for posterior inference. The 90% credible intervals were used to quantify uncertainty in posterior inference. For this paper, μ=300 CFUs/m^2 reflects the average bioburden density

requirement provided by NASA for Mars-bound Planetary Protection Category IVa bioburden sensitive missions.

The dependence of Bayesian inference on prior selection has always been a sticking point between frequentist and Bayesian schools of statistical thought. From the frequentist point of view, Bayesian analysis is biased unless it deploys a "correct" prior; hence, it is vulnerable to biased results. On the other hand, it has been known since the 1960s [12] that EB can achieve smaller total squared error risk than a maximum likelihood estimator (MLE). This superior performance justified the efforts for Bayesian prior selection and resulted in numerous publications in this field [13, 14, 15, 16, 17, 18]. The EB techniques are broadly categorized [18] as parametric EB techniques and nonparametric EB techniques. Due to the physical constraints when estimating the bioburden of spacecraft components, including the limitation of the sampleable surface area and the scarcity of positive CFU counts as the usage of EB techniques—particularly the application of informative priors—allows PP engineers to gain increased confidence in the estimation of the true bioburden by pooling more data than could be obtained strictly from the surface of interest and applying it to estimate the true bioburden of various spacecraft surfaces.

4. GAMMA-POISSON COMPOUND DISTRIBUTION MODEL

The core of Bayesian inference is the Bayes formula, which inverts information contained in a data set into the estimation of a parameter:

$$\pi(\lambda / x,\alpha) = \frac{L(x/\lambda)\pi(\lambda/\alpha)}{\int_{\Theta} L(x/\lambda)\pi(\lambda/\alpha)d\lambda} \qquad (15)$$

where $\pi(\lambda/x,\alpha)$ is the posterior distribution of the parameter conditioned on a current data set x and hyperparameter α, which defines the prior distribution of $\pi(\lambda/\alpha)$. $L(x/\lambda)$ is a likelihood function that specifies the probability for the given data set x to occur conditioned on the parameter.

Bayesian predictions can be based on both the posterior and prior distributions of the parameter. Instrumental to performing Bayesian prediction is the likelihood of a future data set z, which is defined as $L(z/\lambda)$. This likelihood assesses the plausibility for data z to occur in future hypothetical experiments given a value of the parameter λ. Combining this likelihood with a prior distribution on the parameter, we get what is called a prior predictive distribution:

$$\pi(z/\alpha) = \int_{\Theta} L(z/\lambda)\pi(\lambda/\alpha)d\theta \qquad (16)$$

Prior predictive distribution is an example of a compound distribution model defined as the data distribution marginalized over parameters of the mixing distribution. The compound distribution model, referenced in Eq. 16 has been the backbone of empirical Bayesian methods for decades.

The integral in Ref. 16 appears in the denominator of the Bayes theorem and is known under multiple names such as evidence, marginal data distribution, and prior predictive distribution. The compound distribution model is usually used in EB methods to obtain an estimate of the hyperparameter α-$\hat{\alpha}$, which can be used in Eq. 15 in lieu of the true value of α to specify the prior in Bayesian inference:

$$\pi(\lambda / x,\alpha) = \frac{L(x/\lambda)\pi(\lambda/\hat{\alpha})}{\int_{\Theta} L(x/\lambda)\pi(\lambda/\hat{\alpha})d\lambda} \qquad (17)$$

In this paper, a Gamma-Poisson compound distribution model is used:

$$\mathrm{NB}\left(x;\alpha,\underbrace{\frac{\beta}{\beta+e}}_{p}\right) = \int_0^{\infty} \underbrace{\frac{(\lambda\cdot e)^x}{x!}e^{-\lambda\cdot e}}_{Likelihood} \cdot \underbrace{\frac{\beta^{\alpha}\cdot\lambda^{\alpha-1}\cdot e^{-\lambda\cdot\beta}}{\Gamma(\alpha)}}_{Prior} d\lambda \qquad (18)$$

where NB is a negative binomial distribution with parameters α and p, the prior is a Gamma distribution with parameters α and β, x-CFU is the count, e is the exposure, and the likelihood is Poisson. Since the compound distribution for the Gamma-Poisson is NB, its empirical moments can be equated with theoretical ones and the values of hyperparameters α and β can be estimated to be used in Bayesian inference [19]. This technique is called the method of moments (MOM) and was used in this paper for EB. A schematic representation of the Gamma-Poisson model and its application in PP PRA is illustrated in Figure 1, where the unknown prior distribution is assumed to be Gamma, which yields an unobservable realization of parameters λ_1, λ_2,…, λ_k, each representing component-specific bioburden density. Here the subscript k is the number of sampled components. Further, each λ_i, i={1,2,…,K} produces an observable number of CFU x_i for a given exposure e_i, according to a known probability distribution, in this case Poisson, x_i~Poisson (e_i,λ_i). We are interested in recovering the prior density Gamma parametrically having only the sample of observed NB variates x_1, x_2,…, x_M and known exposures e_1, e_2,..., e_k. In summary, by using the EB approach we are going backward from observed empirical data to a plausible prior distribution, which is subsequently used in Bayesian inference. The parameters of the two prior distributions used in this paper are summarized in Table 4.

Table 4. Parameters of Gamma distribution for two priors

Prior distribution	α- shape	β - rate	Mean	Variance
CNI	0.5	0.0017	300	1.8e+05

MOM	0.1657	0.0011	144.1434	1.2539e+05

While the rate parameters are almost identical for two distributions, the shape parameters are different, which affects the MOM of the prior distributions. For example, the mean value for MOM is more than twice smaller, while the variance is about 30% smaller. The smaller variance of the prior distribution obtained with MOM will cause posterior inference to rely more on prior information. The means of the posterior and predictive distributions have been used as posterior summary statistics. The 90% credible intervals were used to quantify uncertainty in posterior inference. For this paper, μ=300 CFUs/m^2 to reflect the bioburden density requirement provided by NASA (NPR 8020.12). Credible intervals were calculated as an inverse of Gamma distribution and NB distribution function [15]. Bayesian model selection has been performed using the Bayes factor (BF) approach, which relies on the comparison of posterior odds for different models, thereby selecting a model that is most supported by the observed data. If there are two competing models represented by two different prior distributions parameterized as Gamma (α1, β1) and Gamma (α2, β2) then for each model, the marginal data likelihood is calculated as:

$$P(X/\alpha_i,\beta_i) = \int_0^{\infty} \underbrace{\frac{(\lambda\cdot E)^X}{X!}e^{-\lambda\cdot E}}_{Likelihood} \cdot \underbrace{\frac{\beta_i{}^{\alpha_i}\cdot\lambda^{\alpha_i-1}\cdot e^{-\lambda\cdot\beta_i}}{\Gamma(\alpha_i)}}_{Prior} d\lambda, i = 1,2 \qquad (19)$$

The BF is then calculated as:

$$F = \frac{P(X/\alpha_1,\beta_1)}{P(X/\alpha_2,\beta_2)} \qquad (20)$$

which is the ratio of the total probability of observed data to occur under the two different models.

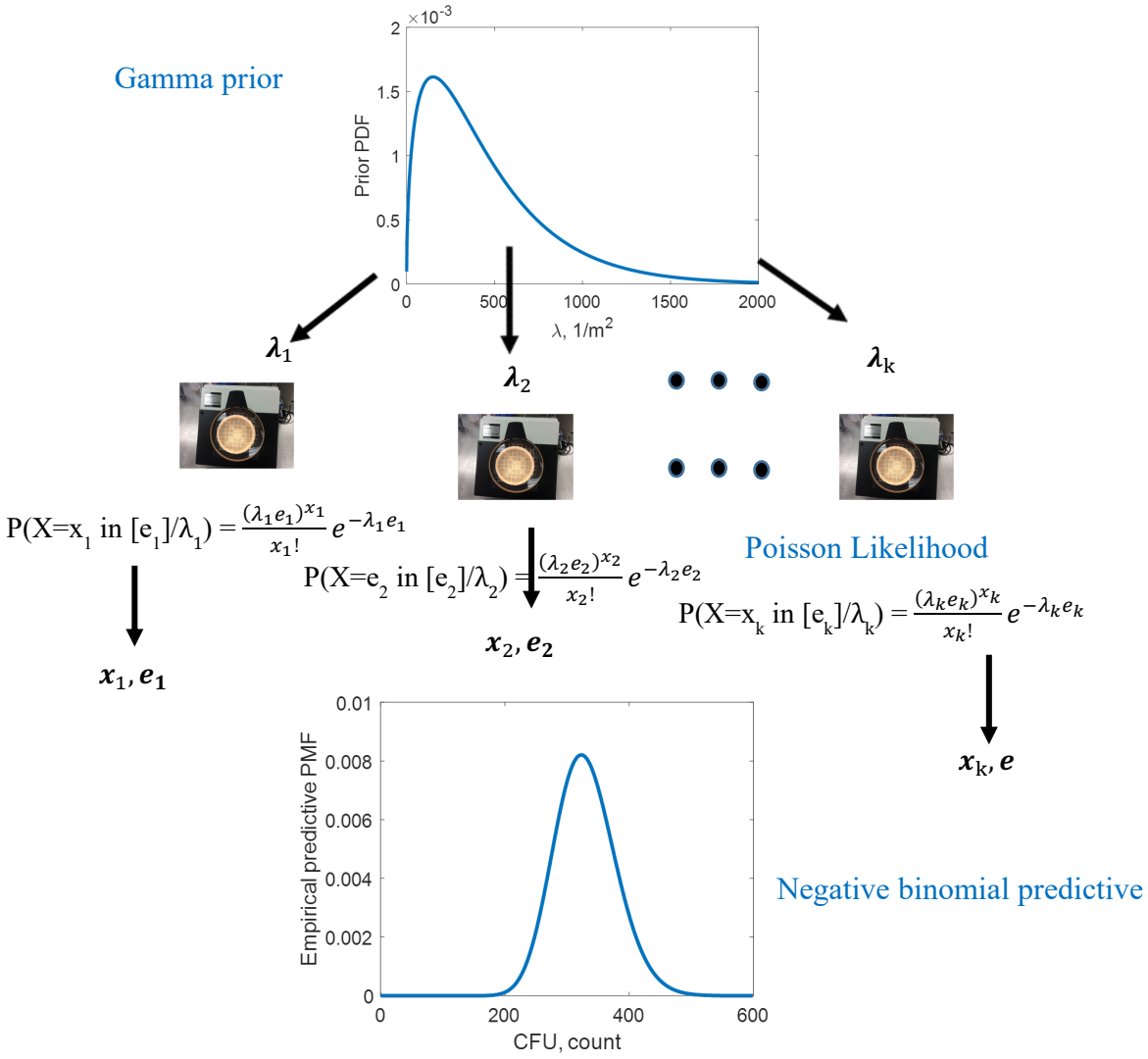


Figure 1. Gamma-Poisson Bioburden Compound Distribution Model

5. RESULTS AND DISCUSSION

The analysis of Tables 5-12 reveals that on the average, MOM produces lower values of estimated bioburden density and total CFU counts. For example, for component 9, the MOM posterior mean is 3 times lower than for CNI, while the predictive mean is more than 3 times smaller. Also, MOM produces tighter credible intervals than CNI. These observations can be explained by a lower mean value of prior distribution for MOM than for CNI and also smaller MOM variance. The difference between the two prior distributions is especially pronounced for the components with small counts of CFU. As the number of counts grows, the difference between the two prior distributions diminishes as the likelihood starts to dominate. The uncertainty in posterior inference is lower if a larger area is sampled, as evidenced from components 283 and 243, for example. The posterior credible intervals are tighter for 283 than for 243. However, for predictive inference, the situation is reversed as a higher percentage of component 243 was sampled; hence, we can become more confident in forecasting the total bioburden. For the large number of counts, the results of applying both priors are nearly identical, indicating that in this case the data overcome prior assumptions and dominates the inference. The mean values of predictive distribution are correlated with the total CFU count found on the component.

It is also instructive to compare components 73 and 300, which have nearly identical sampled areas, but the CFU counts differ only by one. As expected, since component 300 has a higher CFU count, its posterior mean values are higher for both priors, while its posterior credible intervals are wider, reflecting higher variability in the data. Component 300 has much wider predictive credible intervals as a lower percentage of its total area has been sampled. In general, the uncertainty in the posterior estimate of the bioburden density depends on the number of counts and sampled area, while the uncertainty in the predictive inference depends on sampling percentage. If the number of counts is the same, the component with a larger sampling area will have a smaller bioburden uncertainty, while the component with a larger sampling percentage will have a smaller predictive uncertainty. This observation can be used for cost-benefit optimization during mission planning to optimize the number of samples required and help minimize the number of cleaning and resampling sessions.

Components 38 and 261 with the largest number of counts (12 and 52) produced results with the largest uncertainty for both posterior and predictive inference. This reflects the large variability in sampling data for those components. This observation suggests that components with a larger number of counts will benefit in terms of estimation accuracy from applying more restrictive informative priors to increase the effect of regularization.

For comparison, Table 13 shows bioburden density values obtained using NASA's legacy and current approach. The legacy 3-σ approach is a purely frequentist approach based on the assumption of Gaussian statistics for CFU distribution and estimation of the mean value of this distribution and its standard deviation, σ [6]. Having estimated the mean value, the 3σ is added to it for conservatism. As can be seen from Table 13, this approach systematically produced bioburden values higher than the Bayesian approach with CNI prior. On the other hand, the current weighted average approach employs an ad-hoc Bayesian procedure equivalent to using a uniform prior. In contrast to the uniform prior approach used in this paper, the weighted average approach uses one more parameter in the calculation of the effective area sampled-sampling efficiency. Since sampling efficiency is not a directly observable value, it is an estimate, so it had been chosen not to use it in the Bayesian approach as described in this paper. While the 3-σ approach envisages the calculation of confidence intervals and the weighted average does not, neither technique actually currently reports them for requirements compliance. In comparison to Bayesian analysis, both techniques generally produce higher values of bioburden density, thus ultimately producing a higher probability of planetary contamination when used in probabilistic risk assessment models.

Table 14 shows BF for comparison of CNI and MOM. In this study, MOM was used as a null hypothesis when compared to CNI. It can be seen that for components with zero CFU counts, there is some evidence in favor of MOM; however, as the number of counts increases, the difference between the two models diminishes. This once again emphasizes the reliance on the data rather than on the prior for the components with larger counts.

Further model development should include the ability to utilize specification values (i.e., when hardware cannot be sampled) and direct hardware test data in a particular distribution.

Table 5. Summary of posterior and predictive inference for component 9.

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m^2	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	2.2889	0.0090	8.7928	1.7350	0	7
MOM	0.7603	4.1117e-08	4.0993	0.5763	0	3

Table 6. Summary of posterior and predictive inference for component 73.

Prior distribution	Posterior mean. Bioburden density	5th percentile of posterior	95th percentile of posterior	Predictive mean, CFU	5th percentile of predictive	95th percentile of predictive

	– λ, CFU/m²	distribution	distribution		distribution	distribution
CNI	0.8094	0.0031	3.1096	2.2180	0	9
MOM	0.2684	1.4519e-08	1.4475	0.7356	0	4

Table 7. Summary of posterior and predictive inference for component 300.

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	2.2315	0.2617	5.8130	11.1579	1	30
MOM	1.7355	0.1267	4.9272	8.6778	0	26

Table 8. Summary of posterior and predictive inference for component 169.

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	7.7452	0.9083	20.1757	4.5309	0	13
MOM	6.0352	0.4408	17.1337	3.5305	0	11

Table 9. Summary of posterior and predictive inference for component 283.

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	4.8059	1.9987	8.5961	57.6713	22	105
MOM	4.5158	1.8133	8.2011	54.1903	20	100

Table 10. Summary of posterior and predictive inference for component 243.

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	47.5504	19.7758	85.0510	14.1700	4	27
MOM	44.8607	18.0138	81.4700	13.3685	4	26

Table 11. Summary of posterior and predictive inference for component 38.

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	15.4671	9.0398	23.2949	154.6710	88	236
MOM	15.0630	8.7294	22.7980	150.6308	85	231

Table 12. Summary of posterior and predictive inference for component 261.

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	1057.0469	829.0645	1307.8988	329.7986	253	414
MOM	1061.3672	831.7609	1314.0833	331.1465	254	416

Table 13. Bioburden densities calculated using InSight-based weighted average and MSL-based 3-σ approaches for eight components.

Component	9	73	300	169	283	243	38	261
MSL-based 3 sigma Bioburden Density – λ, CFU/m²	13.84	4.87	5.96	20.83	5.17	130.14	52.06	2349.53
InSight-based weighted average. Bioburden density – λ, CFU/m²	27.99	17.36	9.54	33.70	11.11	186.6959	9.66	658.47

Table 14. BF for different components using MOM as null hypothesis.

Component	9	73	300	169	283	243	38	261
BF	4.82	6.83	2.31	1.54	1.50	0.74	1.03	0.42

6. SUMMARY

The performance of parameter estimation methods for PRA applications and compliance with bioburden density values for forward PP is an important technical issue, as well as a regulatory issue. For the numerical estimates produced by bioburden calculations to be taken seriously, they should reflect the bioburden density values observed through test data and provide an estimate of uncertainty for these values. For forward PP, they are a direct reflection of the compliance to the bioburden requirements. Also, the multitude of parameter estimation methods used in PRA raises the question of the "best" or at least the most appropriate method for a given application. Bayesian inference is widely used in modern PRA applications with three major approaches to prior selection, which are all tightly intertwined with one another, including noninformative priors with subsequent updating, EB, and hierarchical Bayes. Each one of these approaches depends on prior selection and elicitation methods. In this paper, we retrospectively analyzed the performance of Bayesian inference with noninformative priors, EB inference, and two additional approaches that have been implemented by previous missions for PP bioburden estimations. Our results indicate that both Bayesian methods produced results comparable with NASA legacy mathematical approaches and currently utilized approaches.. This was likely due to the use of conservative factors, such as the replacement of zero CFU counts with a NASA policy derived CFU count of 1 and the application of sampling device and processing efficiencies. Between the Bayesian techniques, for components with lower CFU counts, the EB prior was favored over the noninformative prior; however, for larger CFU counts, the difference between the two priors is negligible. This demonstrates the importance of prior selection for components with low CFU counts, which normally dominate the collected data sets. Hence, for the majority of PP data with low CFU, EB is recommended as the approach that should be utilized. While both Bayesian approaches provide credible intervals for posterior and predictive inference, they do not account for uncertainty in the hyperparameters of Gamma distribution. This can be provided by hierarchical Bayes, which will be the subject of future work.

In addition to hierarchical Bayes being the subject of future work, model validation will also be conducted. This may include InSight use cases in that a portion of the data from a given sample component used to generate the data-driven prior. Several InSight model validation scenarios could be envisioned to demonstrate the feasibility of the mathematical approach for the application of spacecraft bioburden predictions. Time-trending, similar cleanroom manufacturing environments, and similar microbial reduction processes were the datasets that could be used to cross-validate the models. For the time-trending datasets spacecraft health status data collected from 2013−2016 on the cruise stage solar arrays, the lander primary structure, and the lander multilayer insulation, as well as the instrument tether box groups, which were used from 2016−2018, and the associated raw CFU observed was compared to the model output. Another use case for PP engineering would be the prediction of like-hardware microbial processing so that an allocation could be established for a unique vendor or specific group. To validate the mathematical model for this engineering use case, a telecommunications system was selected from a particular vendor with the InSight values to cross-validate the next missions' identical system and the lander honeycomb structure in that multiple panels were fabricated in the same facility and treated the same, but had different delivery dates. The prediction of microbial reduction processes on surfaces would be an essential use case as over 90% of the surfaces on a spacecraft undergo some form of microbial reduction processing. To predict the bioburden densities of subsequently processed microbially reduced hardware, precision cleaned propulsion lines and alcohol wipe cleaned electronics chassis could be chosen as test cases from the InSight mission. These model validation use cases will be considered as important use cases for future studies.

Implementing a Bayesian statistical approach to perform bioburden density estimations will: (1) facilitate the application of historical datasets and engineering judgement in estimating the total bioburden and bioburden density; (2) assign appropriate confidence intervals and account for uncertainty using a methodological approach; and (3) allow for the prediction of bioburden throughout the lifecycle of a project. Given the increasing complexity and sensitivity of future NASA missions, such as those headed for outer planets and the proposed Mars sample return campaign, a mathematical technique that is documented and vetted by the associated stats and PP scientific communities will serve to allow requirements compliance and consistency in performing PP risk assessments. A Bayesian statistical approach will additionally allow for advanced planning of PP implementation approaches and cost/benefit analysis to be performed in order to optimize the number of samples taken throughout the lifecycle of the project.

Copyright Notice

© 2019 California Institute of Technology. Government sponsorship acknowledged.

U.S. Government work not protected by U.S. copyright.

ACKNOWLEDGEMENTS

The research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. We would like to thank the InSight mission team of Ryan Hendrickson,

Gayane Kazarians, Lisa Guan, Sarah Cruz, and Pat Bevins for providing their support in data generation. The authors would also like to thank Art Avila and Melissa Jones for their management and support of this project.

REFERENCES

Note that the references listed below are publicly available. Non-publicly available internal documents to JPL, NASA, and ESA have also been cited with relevant context detailed in the text.

[1] R. Angelotti, J. L. Wilson, W. Litsky, and W. G. Walter, "Comparative evaluation of the cotton swab and Rodac methods for the recovery of Bacillus subtilis spore contamination from stainless steel surfaces," *Health Lab. Sci.*, vol. 1, pp. 289−296, October 1964.

[2] A. R. Hoffman and D. A. Winterburn, "Microbial burden prediction model for unmanned planetary spacecraft (Final report)," NASA-CR-12746 7, Jet Propulsion Laboratory, Pasadena, CA, USA, 30 June, 1972. [Online]. Available at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720020398.pdf (last accessed 1 October, 2019).

[3] J. R. Puleo, N. D. Fields, S. L. Bergstrom, G. S. Oxborrow, P. D. Stabekis, and R. Koukol. "Microbiological profiles of the Viking spacecraft," *Appl. Environ. Microbiol.*, vol. 33, no. 2, pp. 379−384, February 1977.

[4] J. Barengoltz, "Mars Pathfinder: Planetary protection implementation document," JPL Rep. D-13645. Jet Propulsion Laboratory, Pasadena, CA, USA, 1996.

[5] K. F. Man, C. T. Farguson, and A. R. Hoffman, "Mars Exploration Rover (MER) project environmental assurance program summary report," JPL Rep. D-26957. Jet Propulsion Laboratory, Pasadena, CA, USA, 8 October, 2003.

[6] R. A. Beaudet, "The statistical treatment implemented to obtain the planetary protection bioburdens for the Mars Science Laboratory mission," *Adv. Space Research.*, vol. 51, no. 12, pp. 2261−2268, December 2013.

[7] R. Hendrickson, J. N. Benardini, F. Chen, G. Kazarians, J. Willis, J. Witte, P. Vaishampayan, S. Cruz, P. Bevins, L. Guan, A. Seuylemezian, and B. Shirley, "InSight planetary protection status," 42nd COSPAR Scientific Assembly, Pasadena, CA, USA, 14−22 July 2018.

[8] National Aeronautics and Space Administration, *Planetary Protection Provisions for Robotic Extraterrestrial Missions*, NPR 8020.12D, Washington DC, USA, 20 April 2011.

[9] M. DiNicola, K. McCoy, C. Everline, K. Reinholtz, and E. Post, "A mathematical model for assessing the probability of contaminating Europa," 2018 IEEE Aerospace Conference, Big Sky, MT, USA, pp. 1−20, 3−10 March 2018.

[10] C. L. Atwood, J. L. LaChance, H. F. Martz, D. J. Anderson, M. Englehardt, D. Whitehead, and T. Wheeler, *Handbook of Parameter Estimation for Probabilistic Risk Assessment*, NUREG/CR-6823, SAND2003-3348P, U. S. Nuclear Regulatory Commission, Washington DC, USA, 2003.

[11] C. L. Atwood, "Constrained noninformative priors in risk assessment," *Reliab. Eng. Syst. Safe.*, vol. 53, no. 1, pp. 37−46, 1996.

[12] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Systems Sci. Cybernetics*, vol. 4, no. 3, pp. 227−241, 1968.

[13] W. James and C. Stein, "Estimation with quadratic loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 361–379, University of California Press, Berkeley, CA, USA, 1961.

[14] N. O. Siu and D. Kelly, "Bayesian parameter estimation in probabilistic risk assessment," *Reliab. Eng. Syst. Safe.*, vol. 62, no. 1–2, pp. 89–116, October–November 1998.

[15] H. F. Martz and R. A. Waller, *Bayesian Reliability Analysis*, reprinted with corrections. Krieger Publishing Co., Malabar, FL, USA, 1991.

[16] B. Efron, "Empirical Bayes deconvolution estimates," *Biometrika*, vol. 103, no. 1, pp. 1–20, March 2016.

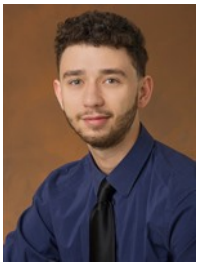
[17] A. Gribok, J. W. Hines, A. Urmanov, and R. E. Uhrig, "Backward specification of prior in Bayesian inference as an inverse problem," *Inverse Probl. Eng.*, vol. 12, no. 3, pp. 263–278, June 2004.

[18] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer, New York, NY, USA, 1985.

[19] C. N. Morris, "Parametric empirical Bayes inference: Theory and application," *J. Amer. Statist. Assoc.*, vol. 78, pp. 47–65 with discussion, March 1983.

BIOGRAPHY

J. Nick Benardini *received a B.S in Microbiology from The University of Arizona in 2002 and a PhD in Microbiology, Molecular Biology and Biochemistry from The University of Idaho in 2008. He has worked as scientist in the field of in environmental microbiology studying extreme environments and has actively participated on the implementation teams for the Mars Exploration Rovers, Mars Science Laboratory, InSight 2018 and Mars 2020 Missions. He has worked at the Jet Propulsion Laboratory for 12 years and is currently one of NASA's Subject Matter Experts for PP and the Forward Planetary Protection lead for the Mars Sample Return Campaign.*



Arman Seuylemezian *received a B.S. in Biology from California State Polytechnic University, Pomona and is pursuing a Master's degree in Bioinformatics from Johns Hopkins University. He has taken roles in the Biotechnology and Planetary Protection Group, characterizing, identifying, and archiving bacterial isolates from spacecraft associated surfaces and cleanroom environments. He has also supported flight project implementation for the InSight and Mars 2020 missions. He is interested in applying new and emerging computational biology technique and technologies to better assess microbial bioburden, present on spacecraft associated surfaces.*

Andrei Gribok *is a Level 4 Research Scientist at Idaho National Laboratory, Department of Human Factors, Controls, and Statistics. He received his Ph.D. in Physics from Moscow Institute of Biological Physics in 1996 and his B.S. and M.S. degrees in systems science/nuclear engineering from the Moscow Institute of Physics and Engineering in 1987. Dr. Gribok worked as an instrumentation and control researcher at the Institute of Physics and Power Engineering, Russia, where he conducted research on advanced data-driven algorithms for fault detection and prognostics. He also worked as an invited research scientist at Cadarache Nuclear Research Center, France, where his research focus was on ultrasonic visualization systems for liquid metal reactors. Dr. Gribok holds the position of Research Associate Professor with the Department of Nuclear Engineering, University of Tennessee, Knoxville, TN. Dr. Gribok was employed as a Research Scientist with the Telemedicine and Advanced Technology Research Center of the U.S. Army Medical Research and Materiel Command from 2005 to 2015. Dr. Gribok has also served as a member of a number of international programs including the IAEA coordinated research program on acoustical signal processing for the detection of sodium boiling or sodium-water reaction in LMFRs and large-scale experiments on acoustical water-in-sodium leak detection in LMFBR. Dr. Gribok is an author and co-author of three book chapters, over 40 journal peer-reviewed papers, and numerous peer-reviewed conference papers. He is also a co-author of the book, "Optimization Techniques in Computer Vision: Ill-Posed Problems and Regularization," published by Springer in 2016.*